Choice Models in Operations

Lecture 7 : NL Model (continued) and ML Model

Instructor: Srikanth Jagabathula *Scribe: Ziran Liu*

Setting: We have m nests and n products in each nest.

$$U_{ij} = W_l + Y_{lj} + \tilde{\varepsilon}_l + \tilde{\varepsilon}_{lj},$$

where l indicates the nest, j indicates the product, W_l and Y_{lj} are deterministic, $\tilde{\varepsilon}_l$ and $\tilde{\varepsilon}_{lj}$ are random.

Objective: Capture any correlations in utilities among the products. (eg. Red Bus/Blue Bus)

Assumptions:

(A1) $\tilde{\varepsilon}_l$ and $\tilde{\varepsilon}_{lj}$ are ind.

(A2) $\tilde{\varepsilon}_{lj} \overset{D}{\sim} Gumbel(0, \mu_l)$, $\mu_l > 0$.

(A3) [Key Assumption] $\tilde{\varepsilon}_l$'s are distributed s.t. $\max_{j \in S_{l'}} U_{l'j} \overset{D}{\sim} Gumbel$ with scale $\lambda_{l'}$ for all subset $S_{l'} \subseteq S_l$.

With the above assumptions, we can show that

$$\begin{aligned} &\Pr(jl \text{ when all items are offered}) \\ = \quad & \frac{e^{\lambda_l W_l + \frac{\lambda_l}{\mu_l} IV_l}}{\sum_{l'=1}^{m} e^{\lambda_{l'} W_{l'} + \frac{\lambda_{l'}}{\mu_{l'}} IV_{l'}}} \cdot \frac{e^{\mu_l Y_{lj}}}{\sum_{j' \in S_l} e^{\mu_l Y_{lj'}}}, \end{aligned}$$

where $IV_l := \log(\sum_{j \in S_l} e^{\mu_l Y_{lj}})$ is Inclusive Value of nest l. $\frac{1}{\mu}$ is the location parameter

1 Correlation Structure

1.1 Correlation among products in the same nest

Recall that $\varepsilon_{lj} = \tilde{\varepsilon}_l + \tilde{\varepsilon}_{lj}$. Then

$$\mathrm{Cov}(\varepsilon_{lj}, \varepsilon_{lj'}) = \mathrm{Cov}(\tilde{\varepsilon}_l + \tilde{\varepsilon}_{lj}, \tilde{\varepsilon}_l + \tilde{\varepsilon}_{lj'}) = \mathrm{Var}(\tilde{\varepsilon}_l), \text{ [From (A1)]}$$

$$\mathrm{Var}(\tilde{\varepsilon}_{lj}) = \mathrm{Var}(\tilde{\varepsilon}_l + \tilde{\varepsilon}_{lj}) = \mathrm{Var}(\tilde{\varepsilon}_l) + \mathrm{Var}(\tilde{\varepsilon}_{lj}) = \mathrm{Var}(\tilde{\varepsilon}_l) + \frac{\pi^2}{6\mu_l^2}, \text{ [From (A1)]}$$

and

$$\mathrm{Var}(\tilde{\varepsilon}_l) = \mathrm{Var}(\varepsilon_{lj}) - \mathrm{Var}(\tilde{\varepsilon}_{lj}) = \frac{\pi^2}{6\mu_l^2}\left[1 - \frac{\lambda_l^2}{\mu_l^2}\right].$$

Since $\text{Var}(\cdot)$ is ≥ 0, we have that $0 \leq \lambda_l \leq \mu_l$. Also, we have that the utilities of the items in the same nest are always positively correlated.

$$\text{Cov}(\varepsilon_{lj}, \varepsilon_{l'j'}) = \text{Cov}(\tilde{\varepsilon}_l + \tilde{\varepsilon}_{lj}, \tilde{\varepsilon}_{l'} + \tilde{\varepsilon}_{l'j'}) = 0.$$

1.2 Correlations between items across nests

NO correlations between items from different nests.

$$\begin{aligned}
\text{Cor}(\varepsilon_{lj}, \varepsilon_{lj'}) &= \frac{\text{Cov}(\varepsilon_{lj}, \varepsilon_{lj'})}{\sqrt{\text{Var}(\varepsilon_{lj})\text{Var}(\varepsilon_{lj'})}} \\
&= \frac{\frac{\pi^2}{6\mu_l^2}\left[1 - \frac{\lambda_l^2}{\mu_l^2}\right]}{\sqrt{\left(\frac{\pi^2}{6\lambda_l^2}\right)^2}} \\
&= 1 - \frac{\lambda_l^2}{\mu_l^2}.
\end{aligned}$$

We can think λ_l/μ_l as a measure of the degree of the correlation among the products in the same nest, $0 \leq \lambda_l/\mu_l \leq 1$.

1.3 Normalization of the Nested Logit model

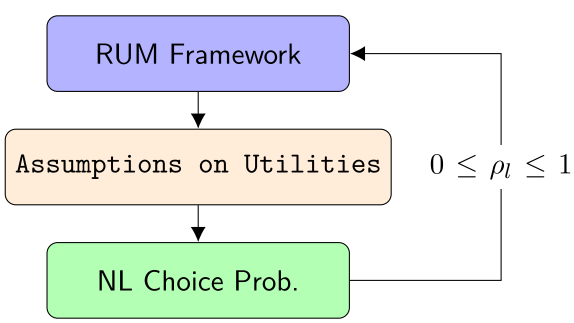
It is possible to estimate both λ and μ just from choice data.

We can only estimate that ration $\frac{\lambda_l}{\mu_l}$ for each nest. Therefore, we normalized and set $\lambda_l = 1$ for all l.

And, let the *nest disimilarity parameter* $\rho_l = \frac{1}{\mu_l}$ and $V_{lj} = e^{W_l} + \mu_l Y_{lj}$. Then, choice probability

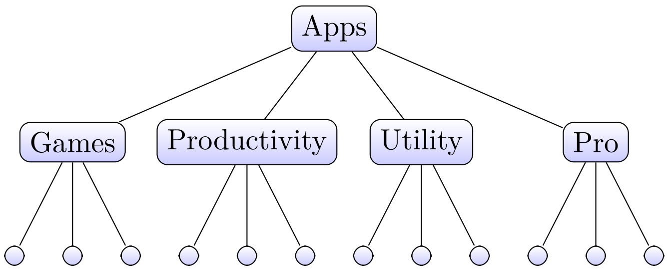
$$\Pr = \frac{(\sum_{j' \in S_l} V_{lj})^{\rho_l}}{\sum_{l'=1}^{m} (\sum_{j' \in S_l} V_{l'j'})^{\rho_{l'}}} \cdot \frac{V_{lj}}{\sum_{j' \in S_l} V_{lj'}}.$$

What happens when $\rho_l > 1$ for some l? We still get a valid stochastic choice rule that obeys the usual probability laws. However, *the resulting stochastic rule is not rational.*

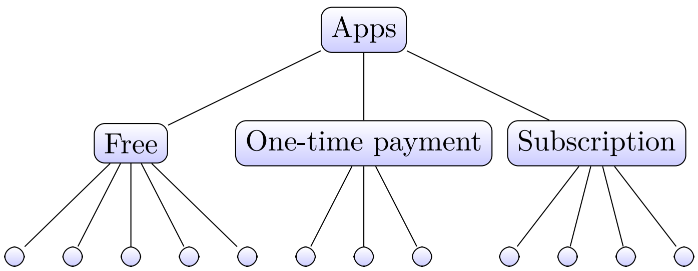


1.4 Examples of applications

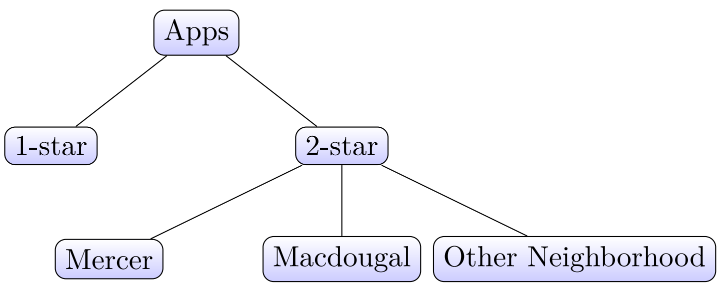
1. Model the purchase of apps on the apps on the App Store.



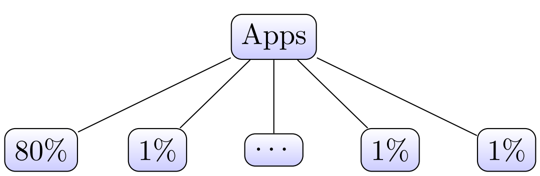
1'. Another way of modelling the purchase of apps on the apps on the App Store .



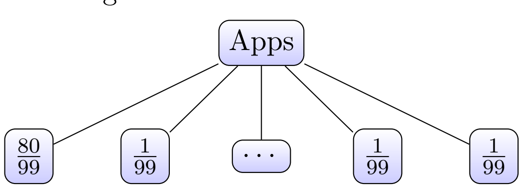
2. Decision tree of finding a hotel.



3. Modelling the purchase option.

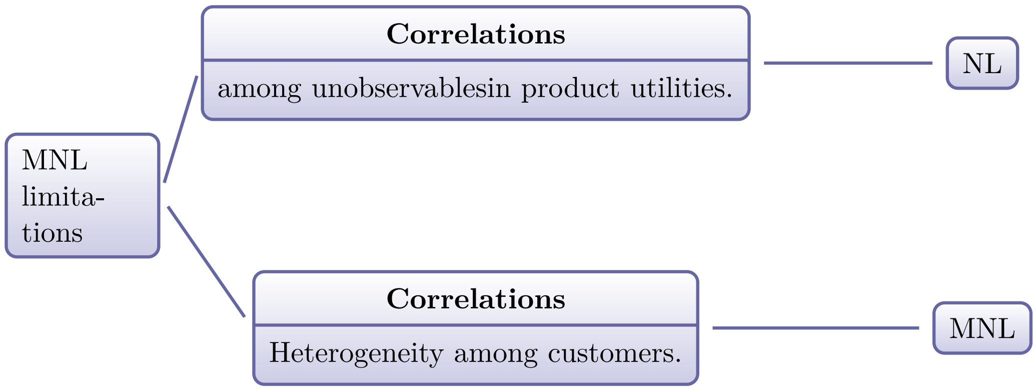


Furthermroe, if we cut off the last 1% population, the purchase option becomes the follwoing.



2 Mixed Logit Model (ML)

Let $U_j = V_j + \varepsilon_j$, for $j = 1, \dots, n$. Objective: To capture customer-level heterogeneity. In the process, we introduce correlations among product utilities.



Consider customer i and j,

$$\begin{aligned} V_{ij} &= f(\text{attributes of customer } i \text{ and product} j) \\ &= \sum_k \beta_{ik} \cdot n_{ijk} = \beta_i^T \cdot x_{ij}. \end{aligned}$$

We suppose that each customer i makes choices according to an MNL model. In particular,

$$\mathbb{P}(j|S) = \frac{e^{\beta_i^T} x_{ij}}{\sum_{j' \in S} e^{\beta_i^T \cdot x_{ij'}}}.$$

We let each customer have a different β_i, which are also referred to as the "taste" vectors of the customers. In other words, customers have different "tastes". β_i's are also called the utility partworth.

2.1 Example

Two attributes: Color (Black, Red) and Size (Small, Large). Full factorial product universe: all possible attribute-level combinations.

$$\begin{aligned} BS &= \beta_B + \beta_S \\ BL &= \beta_B + \beta_L \\ RS &= \beta_R + \beta_S \\ RL &= \beta_R + \beta_L \end{aligned}$$

If there is sufficient data per customer, then we can describe the purchase/choice behavior of each customer using a separate MNL model. However, in many partial settings, there are only a few samples available for each customer. Therefore, a very general model with a separate β_i for each customer will not be identifiable. To address this issue, we make the model parsimonious in the number of parameters by making distributional assumptions on β_i's.

There are two ways in which we can make distributional assumptions.

I. <u>Discrete</u>: We assume that the underlying customer population is compromised of K classes/segments s.t. all the customers in class k have teh parameter/taste vector β_k. we also assume that class k has size $\alpha_k \geq 0$ such that $\sum_{k=1}^{K} \alpha_k = 1$.

This results in the following generative model when teh customers class membership is not identified. In each choice instance, the customer samples her class membership s.t. class k is chosen with probability α_k. Conditioned on choosing an item according to an MNL model with parameter vector β_k.

$$\mathbb{P}(j|S) = \sum_{k=1}^{K} \alpha_k \cdot \frac{e^{\beta_k^T \cdot x_j}}{\sum_{j' \in S} e^{\beta_k^T \cdot x_{j'}}}.$$

Let's consider the following cases:

1. Suppose we are given aggregated sales transactions data. In particular, for a collection of subsets$S_1, S_2, \dots, S_m$, we are given n_{jt} = the number of purchases of item j when S_t was offered, $\forall j \in S_t, t = 1, \dots m$. The date are aggregated all customers and we don't observe the customer ID. These data are also called aggregated transaction data. Assuming that customers make choices according to a K-LC-MNL model. The data likelihood is

$$\prod_{t=1}^{m} \prod_{j \in S_t} \left(\mathbb{P}(j|S_t)\right)^{n_{jt}} = \prod_{t=1}^{m} \prod_{j \in S_t} \left(\sum_{k=1}^{K} \alpha_k \frac{e^{\beta_k^T \cdot x_j}}{\sum_{j' \in S_t} e^{\beta_k^T \cdot x_{j'}}} \right)^{n_{jt}}.$$

2. Suppose, instead, we are given panel data, i.e., the transactions are tagged by the customer ID. In particular, for $u = 1, \dots, U$, we observe the index of users/customers. Collection of tuples $D_u = (j_{u_1}, S_{u_1}), \dots, (j_{u_n}, S_{u_m})$ s.t. user u purchased item j_{ut} when offered subset $S_{ut}, t = 1, \dots, m$. Assuming that customers purchase according to a K-LC-MNL model, what is log-likelihood

$$\sum_{u=1}^{U} \log \sum_{k=1}^{K} \alpha_k \prod_{t=1}^{m} \frac{e^{\beta_k^T} \cdot x_{j_{ut}}}{\sum\limits_{j'_{ut} \in S_{ut}} e^{\beta_k^T \cdot x_{j'_{ut}}}}.$$